Trust me I'm a computer?

Harold Thimbleby HonFRCP • www.harold.thimbleby.net

Abstract. Computers, IT, digitisation, apps — whatever we call it — is everywhere in healthcare, and it is also racing ahead of healthcare and creating dreams and exciting opportunities for quality improvement and exciting transformation. We want a paperless NHS. Yet we have to be careful what we wish for. We are most familiar with consumer IT, our own personal phones and tablets, but our enthusiasm for this must not be confused with what might be best for healthcare.

I use an Apple laptop, an iPad and an iPhone. They are brilliant, but a month ago my four year old laptop broke, and I was thrown back onto relying on just my iPad and iPhone on their own.

I discovered something very interesting: the iPad is *not* brilliant. In fact, it is really irritating and slows down almost everything I try to do.

Previously, when I had both a laptop and an iPad, I used each for what they were good at, and they are very impressive at what they are good at!

Reading stuff is very easy on the iPad — it is great in bed, in a cramped airplane seat, and even in the bath. The iPad is great for skimming my email and deleting junk, keeping up with social media and so on. The laptop, on the other hand, is great for writing and replying to email. I do a lot of presentations, and the laptop is great for preparing and presenting talks. In short, each device had carved its own distinctive niche in my life.

But when my laptop broke, I stopped having any choice; I *had* to do almost everything on my iPad. Replying to mail, for instance, often means reading the email, looking at my diary, writing the reply, and perhaps finding a spreadsheet or something else that has been asked for. On a laptop this is very easy. On the iPad it is actually very difficult. You can only do one thing at a time, and you cannot see more than one thing at once, so there is a huge amount of work going backwards and forwards between apps. You also need a surprising amount of quirky knowledge to get things done at all: for example, if you want to email a spreadsheet to someone, you can't just start to reply to the email and then attach the spreadsheet, you have to plan ahead and start with the spreadsheet, then write the email. In fact I often start a reply, then notice I need to attach a spreadsheet or other document, so I copy my reply, go to the spreadsheet, create a new email, then paste in the old email (if I haven't lost it en route). It's quite a workaround. There may be a better way to do this, but it isn't obvious, as it was on the laptop.

In the "old days" when my laptop worked, I would have written this article you are reading on it. Instead, I had to write it on my iPad (using Apple's Pages, which is an app like Microsoft Word). Since I have an iPad keyboard, I used the iPad in landscape mode, and I happily started to write this article. Then I noticed that the article was in landscape but I really needed it in portrait for this journal. It turns out you cannot change the paper orientation: the Pages help for the iPad says if you need to change the orientation, do so on a mac — but I don't have one! So previously, when I used Pages a bit on my iPad it seemed wonderful, but when I really used it, it turns out to have major limitations. Fortunately, I could use a work-around — I copied the entire document, created a new one (in portrait), then I pasted the entire document back into the new document. Unfortunately, the copying does not copy everything, so I had a bit of tidying up to do, like restoring page numbers that had got mangled.

This simple example is typical of much use of IT: it works, but only when we do the right work-arounds to overcome "little" problems.

A few other things happened. We acquired a cat. I wanted to take a photograph of it doing one of its mad acrobatics — the sort of thing to send to the family. I got the iPad out, found the

camera, ... and the iPad asked me if I wanted to upgrade to a new version of the operating system, Apple's iOS. No I don't — I want to photograph my cat! If I don't want to upgrade the iPad, it still asks for my password and asks whether I'd like to upgrade tonight, if not now. Frankly, I don't care. And now the cat has moved, and the cute photo opportunity is lost.

I have just described a problem well known to experimental scientists. **Success bias** (also known as survivor bias) is the problem that you pay too much attention to success and ignore failure (Smith, 2016). A classic healthcare example would be you trial a drug on 100 patients, and then give them a questionnaire on how their life has improved since taking the drug. Well, 100% of respondents say their symptoms have gone. It seems to be an amazing drug. *No*. The problem is that the patients who the drug killed are not going to respond to your survey; they can't. It is inevitable that the only people who respond must be sufficiently healthy to answer the questions.

Your experiment design suffers from success bias.

It is interesting that while we generally recognise success bias as a problem for scientific experiments, we do not notice its effect in our everyday lives.

In my normal life, I used to suffer from success bias with iPads. They are fantastic … *when* they are fantastic. And when they are awkward, I used my laptop instead. The success bias is that I did not notice this "problem" because I can always get on and do what I want to do, and when I use the iPad it is actually amazing, and when I use my laptop it, too, is amazing. But as my story shows, when I have to use my iPad for everything, then I quickly discover sometimes it is not so good; it's even obstructive. Unwittingly, my opinion of my iPad was influenced by success bias.

We risk making a similar mistake with our opinions about IT and healthcare.

We all like to think handheld devices — like iPhones and iPads — are wonderful, so we think that they must also be wonderful for using in hospitals. Indeed, many hospitals have bought lots of iPads. It is obvious, indeed it seems self-evident, that they are wonderful, so giving everyone an iPad must be a really good idea. In fact, everyone wants one anyway, so we will also be making everyone very happy. And iPads are cute and modern, they use wifi and clouds, so we have made everyone happy and modernised our IT at the same time. It must be a no-brainer to do this.

No.

Our ideas about how wonderful modernising to iPads have been influenced by success bias. Just because our experience is that we like iPads, does not mean we have good evidence for using them in hospitals. We probably don't want to photograph cats on the ward, but if an iPad takes as long to use when dealing with a patient, it could lead to unnecessary harm.

Worse, success bias is not the only problem.

IPads were designed to appeal to consumers, and indeed they are very "sexy"; but they were not designed to be used in hospitals. Will they work when they get body fluids on their screen (no); will they work when they are dropped (no); will they work outside of wifi (maybe); will they work throughout a shift (not unless they are fully charged). Nevertheless, as a personal consumer, I think iPads are wonderful — but as a worker trying to do a job, how good are they really?

In our private lives, we can always find work-arounds (like, having a cup of coffee while the software is upgraded) — but at work in a hospital, this is not an option. Our personal success bias has misled us about how effective IT is going to be in hospitals, where delays matter and relaxing while we wait for the IT to reboot is not a realistic option.

Attribute substitution is one of our human tendencies to take short-cuts in decision making (Kahneman, 2012). Deciding whether iPads are right for a hospital is a very complicated decision, which should obviously be evidence-based — how do we do this, where do we find the experiments, what are the lifetime benefits to patient outcomes... and so on? Such a rigorous approach would be dauntingly complicated. Instead, we bought an iPad for our own personal use and it seems truly wonderful (thanks to our unnoticed success bias), so if it is wonderful, why go to the trouble of doing all those tedious experiments when we already know it is so good?

In other words, we substitute *our personal desire* for iPads for the hard work of doing a *rigorous evaluation*. Attribute substitution happens without our being aware of it. We may further rationalise our decision. One of the popular rationalisations not to do rigorous evaluation of IT is that a proper randomised controlled trial (RCT) would take a long time, by which time IT will have moved on so the RCT will be obsolete. So there is no point doing an RCT…

But this argument implicitly assumes IT is wonderful, so an RCT is going to be a waste of time anyway. If I want an iPad, I want one now! Why wait for an RCT to tell us what we already know: we are already certain we know iPads are wonderful.

I hope the problems of success bias and attribute substitution make the problems clear, but let's illustrate the fallacies in a powerful way. If we do an RCT to see whether thalidomide helps in

morning sickness for pregnancy, it will take about 9 months to perform, by which time we will have missed the opportunity to help some people. As we now all know (Kim & Scialli, 2011) this is a stupid argument. In fact, the thalidomide tragedy's legacy is our deep respect for RCTs: they are essential for evaluating clinical interventions. In the 1950s, the presumed safety of thalidomide meant it was an over the counter drug. Thalidomide was much like today's IT: we presume IT works well, and why test its effectiveness when it is available to everyone over the counter? There are TV adverts telling us to buy more, so they must be right. It is self-evident that it is good if it sells so well!

Yet we all know times when IT has crashed, when our PC needed rebooting or when the wifi signal was not available. When we pause to think about it, then, we know *sometimes* IT fails, sometimes quite catastrophically losing our work or worse. We know some hospitals that have been brought to a standstill because of ransomware, where their IT services have been destroyed by a cybersecurity hack. Sometimes, IT problems will result in not just stopping treatment, but in adverse events and may lead to criminal trials as well. Elsewhere I have described a case where nurses were threatened with prison over what turned out to be a spectacular IT failure (Thimbleby, 2017). Unfortunately, UK law itself is not very helpful: the Criminal Justice Act 2003 explicitly says that IT is presumed to work; this means that if something goes wrong, the presumption in law is that it is your mistake and not the IT system's. (The culture that "IT is wonderful" runs deep.)

When we use IT, we may make mistakes. That's why most IT has a delete key — it allows us to correct mistakes we notice. However, we make mistakes because if we noticed we were making a mistake we would simply avoid making the mistake; the implication is we can make mistakes when using IT *and not always notice*. In particular, we may not notice how the IT itself is making mistakes. In the worst case, we may end up in court charged with problems that may have been caused by poor IT, but which we are completely oblivious to (Thimbleby, 2017 describes a major such example).

It sounds obvious, but if we had been aware the IT was problematic, we would have done something about it. So when we or our colleagues end up in court, everyone was almost certainly unaware of the possible role of IT in the incident that put us there. Worse, IT manufacturers will probably argue that their IT is CE marked, etc, so it follows in law that any problems are the user's fault — yours.

The legal system is not on our side; it is as if it has swallowed the success bias problems enthusiastically! There are more alarming reasons: most people in the legal system do not understand the intricacies of IT, so examining IT carefully can lead to more problems in court.

Unfortunately, problems with IT are complicated to describe, let alone understand (which is another reason why people ignore the problems until it is too late).

Here, then, is a very simple problem to describe, which could affect anyone. This is a very simple example, and it is one that I have written so that you can reproduce the results if you so wish: it is a very real problem. Elsewhere I have argued that it is a typical problem (Thimbleby, 2015) — lots of IT behaves in exactly this sort of unreliable way, and as this simple example proves, IT can cause serious problems that are hard to argue against.

Casio are world leaders in making calculators, and calculators are a very simple sort of IT used throughout healthcare for all sorts of common procedures, such as calculating drug doses.

The Casio HR-150TEC (Casio, undated) is a popular desk-top calculator that can also keep a paper record of any calculation. It is important to remember that we use calculators because we cannot reliably do the maths in our heads, so we are very unlikely to be able to spot any errors — if we could, we wouldn't be using calculators in the first place.

Rather than give a complicated example, let's enter the number 7.5. In practice we would do this as part of a larger calculation, but for this example we can ignore that. Suppose we make a typing mistake, and enter 7..5 by mistake (that is, accidentally typing it with two decimal points). The Casio has a delete key to correct errors, so we should press it twice (i.e., to delete the two keystrokes, .5) and retype the 5. Of course, this should result in 7.5. To spell it out: with the first delete we expect to delete the last keystroke, 5, and with the second delete we expect to delete the accidental extra decimal point: so the display should then be 7., so pressing 5 after these corrections will make it 7.5. After all, this is how delete works everywhere.

In fact, the Casio has a bug: its delete key is defective (this is how the Casio is designed — it is not that it worked once and now is broken, but it was *designed* to be defective). Your correction will result in the number 5 being entered in your calculation, not 7.5. You may not notice this "double" error. Your corrected number is not what you intended, and patient harm may result.

If we had done a more realistic calculation, the consequences could be much worse and harder to spot (but unfortunately harder to explain in a short paper). If you ended up in court from this sort of error, the Casio has kept a paper record, a log, of what you did. This could be used as evidence. Unfortunately, the log clearly shows you entered the number 5. Therefore it appears to be your responsibility.

The log "proves" you entered 5 when in fact you entered 7.5 (albeit after making a correction). The log actually says what the Casio did, not what you did. If the Casio is faulty (as it is), its logs are equally faulty. You can check the example I have described above; until Casio upgrade their HR-150TEC (and that is probably too expensive to countenance), the details above are exactly correct.

The Casio is faulty, but it is not faulty enough to cause frequent problems in normal use. In fact, often when you do a calculation and double check it, if anything is wrong, you think it was your fault and you do it again. In reality, sometimes errors are caused by bugs in the calculator, and they are not your fault.

To put this simple example in perspective, Kimberly Hyatt was a paediatrician who committed suicide after making a simple calculation error (Thimbleby, Lewis & Williams, 2015). Unfortunately, we have no idea how her error happened; it may have been because she was using a defective bit of IT, like the Casio calculator, to do the calculation.

Somehow, consumerism, advertising, even the law and our personal (success biased, attribute substituted) experience have conspired to convince us that IT is self-evidently a fantastic thing — and we must have more of it, and preferably more modern IT too — clouds, wifi, handhelds and so on. The NHS *must* go paperless. The NHS must invest in modernising its IT. It seems only polite to indulge individuals who are convinced that iPads (or whatever they are excited about) are the right solution to any problem.

Yet none of these expensive dreams have been adequately tested, let alone tested with the rigors of an RCT. Where did evidence-based procurement go? Why do we allow clinicians to use their own IT when we would think it an unacceptable risk if they used their own kitchen knives just because they "liked" them?

Going paperless can completely eliminate the symptom of bad processes, namely huge stacks of paper, but it does not change the mess that creates all that information — in fact, it turns a simple problem we all understand into a complex one of finding information in a huge pile we cannot see. Our small screens will make reading the information much less reliable. Other things being equal, we are *more likely* to make errors we do not notice when there is no paper. As with all IT innovation, there is a complex trade-off, and it isn't at all obvious what is safer or faster — without doing a rigorous experiment.

Going paperless is easy and panders to our desire for the latest IT, which will have nice screens and run on handhelds. Which we know we think is wonderful. Going paperless is deceptively easy, but it is palliating the symptom and not curing the disease.

Even if IT is found to be wonderful, will it remain cost-effective, when if we have to keep upgrading our IT with the latest things every year? This year we want iPads, but will we want better things next year? I think buying iPads may well be an improvement, but I think we ought to step back before the superficial arguments for getting iPads this year are recycled in a year's time to get the next self-evidently wonderful IT (wrist watches? implants? faster iPads? sterilisable iPads?) It is certain that while we consumers continue pushing ahead, we need better reasons to upgrade IT in hospitals than just catching up with what we like using at home.

"Trust me I am a computer" (the title of this paper) seems an innocuous claim, and indeed it was the sentiment underlying the recent UK Wachter report (2016) promoting increased investment in NHS IT.

When we look more carefully, though, we see that consumer IT was designed to make us feel that IT is wonderful. More precisely, it was designed to stimulate consumerism, and we as consumers love it. In fact, manufacturers would go out of business if we seriously questioned this; they survive by playing to our attribute substitution, by our susceptibility to success bias, to our cognitive dissonance (Tavris & Aronson, 2007) and to many other weaknesses that advertising plays on. But none of our personal experience is reliable when we have to work out the best sort of IT to trust for future hospitals. The mistake is we over-generalise: personal IT may *feel* wonderful for personal use, which is fine, but we must be much more careful if we think IT is wonderful for anything else, and we must be very cautious if we think it's wonderful for things as complicated — and so unlike our personal IT needs — like anything in the NHS! The paper already cited (Thimbleby, 2017) gives a detailed account of a serious hospital IT problem that escalated to a serious court case *because* everybody assumed the IT was completely trustworthy.

Perhaps this article has sounded one-sided?

Koppel and Gordon (2012) give a balanced overview of the wider issues of healthcare, IT and safety; they show the problems are extensive. Our own paper Thimbleby, Lewis and Williams (2015) gives many detailed examples, as well as some positive suggestions for change, particularly through more careful procurement. We focused on "conventional" IT in this paper, but the problems and dangers apply to all complicated hospital equipment with computers inside (so-called embedded computers). There are numerous analogous problems with infusion pumps, linear accelerators, implants, dialysis machines, ventilators and more. Until we stop to think and get appropriate evidence to justify our plans (as we would for using drugs), the unchecked rush to make everything (for instance) wifi enabled (because it sounds trendy) may have unintended consequences ("side effects" if they were drugs) — from cybercrime risks, to increasing human error because the systems are too complicated in unknown ways. And so on.

We need more rigorous evidence, and more nuanced evidence of IT safety and effectiveness — different sorts of IT may be better or worse at different sorts of clinical task. Currently we just do not know. Don't trust me because I am a computer, trust me because there is a meta-analysis of RCTs that show there is a significant, cost-effective benefit to patients and to staff. We need to get rigorous evidence before we can trust computers in hospitals.

Notes

In the present article I referred to the market leaders Apple and Casio, not because they are particularly problematic, but because they are leading companies and they are leading because we — the consumers — think they make worthwhile products we want to buy lots of. I referred to their popular products because my comments can easily be confirmed because these products are so widely available: the facts of my discussion are reproducible.

I did not have space to discuss software in this brief paper (e.g., medical apps, patient record systems), or medical devices with embedded computers (like infusion pumps, ventilators and linear accelerators), but the pitfalls and problems with these types of IT are identical, though more specialist and therefore harder to describe concisely.

References

Casio Computer Co. (undated). *HR-150TEC User's Guide*, Casio, Tokyo, Japan.

D. Kahneman (2012). *Thinking, fast and slow*, Penguin.

J. H. Kim and A. R. Scialli (2011). "Thalidomide: The Tragedy of Birth Defects and the Effective Treatment of Disease," *Toxicological Sciences*, **122**(1):1-6. DOI: 10.1093/toxsci/kfr088

R. Koppel and S. Gordon, eds. (2012). *First do less harm: Confronting the inconvenient problems of patient safety*, Cornell University Press.

G. Smith (2016). *Standard Deviations: Flawed Assumptions, Tortured Data and Other Ways to Lie with Statistics*, Gerald Duckworth & Co Ltd.

C. Tavris and E. Aronson (2007). *Mistakes were made (but not by me)*, Harcourt Inc.

H. Thimbleby, A. Lewis and J. Williams (2015), "Making Healthcare Safer by Understanding, Designing and Buying Better IT," *Clinical Medicine*, **15**(3):258–262. DOI: 10.7861/clinmedicine.15-3-258

H. Thimbleby (2015). "Safer User Interfaces: A Case Study in Improving Number Entry," *IEEE Transactions on Software Engineering*, **41**(7):711–729. DOI: 10.1109/TSE.2014.2383396

H. Thimbleby (2017). "Cybersecurity problems in a typical hospital (and probably all of them)," *Proceedings of the 25th Safety-Critical Systems Symposium*, M. Parsons and T. Kelly, eds., pp.415–439, 2017.

R. Wachter, ed. (2016). *Making IT work: harnessing the power of health information technology to improve care in England*, Report of the National Advisory Group on Health Information Technology in England, UK Department of Health.